

DRAFT ONLY: PLEASE DO NOT QUOTE

Paper Prepared for the Mexico 2008 World Bank Conference on Performance Budgeting, June 9-10 2008, Mexico City.

PERFORMANCE MEASUREMENT: INDICATING THE WAY TO BETTER PUBLIC SERVICES?

Christopher Hood
www.christopherhood.net

All Souls College, Oxford

Abstract

This paper explores three different applications of quantitative indicators for the performance of public services, namely as targets (threshold standards), rankings (league tables and benchmarking) and intelligence (background information for policy and operational intervention). Arguing that there is little usable theory about when each of these applications of performance indicators will or should be used, it briefly explores three simple theoretical perspectives (couched in terms of dilemmas, variable returns to use, and contingency theory). It argues that even with due attention to the parsimony principle in science, a contingency approach needs to be developed.

1. Introduction: Three Uses of Performance Indicators

Public service performance indicators are by no means a new phenomenon. But they have developed markedly over the past two decades at both national and international level, and they are commonly said to be one of the defining features of the so-called New Public Management.

Much has been written about these developments. Some of the discussion both at popular and scholarly level turns on a debate on the value or otherwise of trying to represent public service performance in numbers at all. On the one hand are those who champion the extension of quantitative performance indicators on the grounds that such indicators represent an important move towards more evidence-based public management and away from 'double-bind management'. On the other side of the debate are those who are suspicious of 'public management by numbers' on the grounds that such numbers tend to involve brutal reductionism that will inevitably fail to capture all the important nuances of performance and cause damage if they are taken seriously.

Beyond that fundamental debate, perhaps the three most common themes in the literature on this subject are: practico-descriptive accounts of the development of such indicators as part of performance management systems in various times and places; social-science assessments of the validity and reliability of the statistical indicators used to represent performance; and analyses of the social dynamics that surround the use of such indicators, including the unintended or unanticipated effects that can result from their application (see for instance Blau 1955). All of these themes are found in the UK's current ESRC research programme on public service performance (www.publicservices.ac.uk).

In spite of such scholarship, however, we still do not have much usable theory about when what kinds of performance indicators will or should be used, or about what unintended effects they are likely to produce in what circumstances. We have nothing equivalent to the sort of 'contingency theory' that developed in the 1960s to qualify earlier one-size-fits-all models of organization that had been the ruling orthodoxy in management and administrative studies in the early twentieth century. We know more about the trees than the wood.

Against that background, this paper briefly compares three broad uses of performance indicators, namely as target systems, ranking systems and 'intelligence' systems. It is not claimed that these three types comprise all possible uses of performance indicators,¹ only that they represent three of the most commonly observable types, and that the way that they do or should relate to one another has hitherto been little explored. Accordingly, the aim of this paper is to identify the different needs those three uses of performance indicators meet and the sort of unintended effects they can produce in different cultural conditions.

¹ For instance Gormley and Weimer (1999: 4) identify the array of applications as what they call 'organizational report cards' (rankings in the language of this paper), US Government Performance and Reporting Act data (broadly, targets in the language of this paper), Benchmarking (a variant of rankings from the analytic viewpoint of this paper), Balanced scorecards (a hybrid of targets and intelligence as considered here), Program evaluations, social indicators and disclosure requirements (all three of which can be broadly considered as forms of intelligence in the language of this paper).

2. Target Systems

Target systems are commonly used in the management of casework handling in public organizations, such as welfare, tax and job placement services. They frequently appear at other levels of management too. Applications range from internationally-agreed targets (such as those applying to emissions or poverty reduction) to national targets, for instance on the management of inflation by central banks. Target systems figured prominently in the Soviet system of economic management from the 1930s to the 1990s, and in the war economies of the 1940s. They formed a key part of Frederick Winslow Taylor's (1911) famous system of 'scientific management', originally devised in the 1890s, because the extent to which observed production exceeds or falls short of the quota or target is what determines each worker's reward.

The essence of a target system is the specification of a performance threshold – normally but not necessarily expressed in numbers - to be attained in a defined time period, typically accompanied by negative feedback systems in the form of sanctions or rewards of some kind.

The advantage of target systems in public services lies in their capacity to concentrate the minds of managers, officials and workers on goals that political leaders consider to be important. They can be considered to be a way of facilitating democratic government, insofar as targets are set by elected politicians on the basis of policy stances adopted in election campaigns. And they can seem an attractive alternative to other methods of managing public organizations, particularly that of controlling such organizations by unspecified but contradictory objectives, with the balance subject to continuous alteration by the top echelons of the political system (see Dunsire 1978).

The three most commonly-noted shortcomings of target systems consist of ratchet effects, threshold effects and output distortions. Ratchet effects denote the incentives for managers or workers to hold actual production below the production-possibility level if they expect target-setters to raise next year's targets in the light of this year's observed production (and expect still to be in office in future years).² These effects have been much discussed in relation to the former Soviet target system, and indeed have been attributed by some to the economic failures of the Soviet system after the ending of Stalin's terror regime in the 1950s (Yavlinski and Braguinski 2000). But those effects can be detected in other target systems as well.

Threshold effects denote the incentives for managers or workers to concentrate their efforts on reaching the minimum levels specified by targets rather than to exceed those levels. W Edwards Deming's (2000) well-known criticism of targets as inimical to continuous improvement in organizational performance is based on threshold effects. The classic example of such effects in public services comes when schools are given targets expressed in terms of numbers of exam passes at a specified grade and teachers concentrate their efforts on the narrow band of students who are on the margin of achieving passes at that grade, while ignoring the more and less able students outside that band.

² . See for example Bain et al 1987; Brown, Miller and Thornton 1994: 93; Kornai 1992

Output distortions denote the incentives for managers or workers to concentrate their efforts on achieving success in the specific measures used for targets at the expense of other important objectives that are not included in the target measures. The classic example is the probably apocryphal Krokodil cartoon (Dobb 1970) in which a Soviet nail factory responds to a target of nail production set in tons by producing a single gigantic nail. In public services, one of the most commonly discussed instances of this kind of problem is the phenomenon of schools and teachers 'teaching to the test' where targets are expressed in terms of specified student test scores, and ignoring other key aspects of education (such as sport, art, general social development).

Some of these effects seem to come in the form of tradeoffs arising from the design of target systems. For example, a system in which every production unit is set a different specific target (as applied to the former Soviet Union and applies in some public sector settings) will be particularly vulnerable to ratchet effects but the threshold effects may be easier to control by the target-setting process. Conversely a system in which every production unit is given the same target (as applied to the targets set by the Blair government for English schools and hospitals after 1998) will be less vulnerable to ratchet effects but correspondingly more vulnerable to threshold effects, in so far as incentives for excellent far-above-target performance are weakened and the incentives for attention to individuals who are not crucial to target attainment are absent.

Target systems have been argued by some (notably Barber 2007) to be better adapted to lifting performance from a very low level to something that is broadly satisfactory than to moving from a broadly satisfactory level to one of outstanding excellence. And the extent to which the three classic problems associated with target systems manifest themselves are likely to depend on the culture and society in which those systems operate. Where there is a common culture and low levels of fear as between the higher directorate and managers and workers, output distortions are less likely to occur. Where managers do not expect to be in office for a long period (as Yavlinski and Braguinsky (2000) claim to have applied for the first twenty years of the Soviet target system), the ratchet effect is also less likely to be pronounced.

3. Ranking Systems

In contrast to target systems, ranking systems involve the use of indicators to compare the performance of different individuals, organizations or institutional units in a league. Like target systems, they have a long history in both business and public sector applications. The basic idea can be traced back at least as far as Jeremy Bentham's 'tabular-comparison' principle for managing public services two hundred years ago (see Hume 1981: 157), ranking systems were developing both at international and national level in the nineteenth century, and there were notable theoretical and practical developments in the twentieth century as well, such as the development of tournament theory in economics (Lazear and Rosen 1981). Ranking systems can be found at both national and international level, and come in various forms, from benchmarking exercises to more elaborate leagues. International ranking systems have notably developed over recent decades for issues of governance and public service performance. Gormley and Weimer (1998) who use the term 'organizational report cards' to describe this phenomenon, see ranking systems as an important way of improving organizational performance.

The advantage of ranking systems is that they attract attention from the media and the public as well as from the players whose performance is being ranked. And that attention can in turn be used to encourage saints, shame sinners, and incentivize all the 'rankees' to maintain or improve their relative position. Compared to target systems, the advantage of ranking systems is that they avoid ratchet effects (because there is normally no advantage to be gained from holding production below the level that is possible in any one time period). They may produce threshold effects if the design of the ranking is such that performance only up to a threshold level factors into the indicators. But if all performance indicators are measured on a continuous scale, threshold effects will not occur in a ranking system.

However, ranking systems are also subject to at least three major and related limitations, namely indeterminacy, volatility and output distortions. Indeterminacy is a form of validity problem which arises when the measuring tools available cannot meaningfully distinguish among the cases to be ranked. When statistical confidence intervals are added in, it is common to find that a substantial number of cases in international or national ranking systems are statistically indistinguishable from one another, creating the so-called 'messy middle' problem (Besançon 2003; Jacobs 2007). If only the extreme ends of the distribution are robustly distinguishable, ranking systems that involve high stakes for small movements in the middle-ranked cases will operate as a form of lottery.

Volatility is a form of reliability problem that typically comes from compositing and weighting when several numbers are put together to create a single overall number in a ranking scheme. Given the multiple dimensions on which the performance of public services can be assessed, the logic of boiling them down to a single overall number is hard to resist – and certainly has not been resisted by the architects of numerous well-known ranking systems, such as the World Health Organization's 2000 much-criticized ranking of the efficiency of health systems across the world and the rankings of hospitals and local authorities introduced by the Blair government for England in the early 2000s. But if the rank order of the units varies substantially if small differences are made in the weightings used to combine the numbers or in the particular basket of performance indicators that are used to make up the ranking, the result can seem arbitrary and liable to be challenged by alternative rankings using different weightings or component units.

Output distortion – altering behaviour such that measured performance improves by means of paying less attention to important aspects of behaviour that are not measured for the purpose of rankings – can also arise with ranking systems. For example, if schools are ranked on their performance on exam passes, the pressure to climb up the rankings league may involve distortions such as 'teaching to the test', as already mentioned, and indeed many critics of the 'league table' approach to managing school performance have argued that such distortions can be serious (see for instance Elmore 2000).

As with target systems, the design of ranking systems can also involve unavoidable tradeoffs. Gormley and Weimer (1999) argue that there can be a tradeoff between the scientific validity of ranking systems and their usability and accessibility to the public at large. An alternative view is to take validity and reliability as necessary but not

sufficient conditions for effective ranking, rather than qualities that can be traded off against accessibility or usability. But even in that case, there may be tradeoffs among different aspects of validity and reliability, as I have argued in a study with colleagues of 14 international governance and public service rankings (Hood *et al.* 2007).

Cultural conditions are also likely to be all-important in affecting whether rankings are introduced and how they operate. In some political cultures (Belgium for example), school and hospital league tables seem to be politically unthinkable even though the information needed to construct them is officially collected, and official school league tables have been abolished under egalitarian pressures in three of the four countries of the UK during the 2000s. But in some cultural conditions, rankings can produce fierce competition and even mutual sabotage as individuals and organizations use tactical and strategic manoeuvres to outpoint their rivals.

4. Intelligence Systems

Strangely, the use of performance indicators as intelligence seems to have had rather less analytic attention than their use in target or ranking systems. By intelligence is meant the use of performance indicators as background information available to inform general policy-making or evaluation or particular administrative interventions, but not used in a pre-ordained way as target or ranking systems. The term perhaps implies an analogy with intelligence-gathering in security services, and in some applications that analogy is fairly precise, with the relevant information restricted to a closed circle and subject to intense debates about interpretation.

The use of performance information as intelligence comes in various forms. One is background information collected by agencies and organizations to help them evaluate the entities they oversee. Current examples include medical episode data collected by health agencies, data on aircraft near-misses (collisions narrowly averted) collected from anonymized reporting to aviation authorities, or the various kinds of performance data collected by the World Bank that do not figure in the Bank's published governance indicators but nevertheless inform its lending decisions. There are plenty of historical examples too. For instance, in the eighteenth century the British East India company seems to have collected a great many numbers about the performance of its various trading stations, but did not use them in explicit rankings because of doubts about the quality of the data (see Misra 1970). Another example is the benchmarking data that organizations collect about themselves, for example in the case of data clubs that involve exchange of performance information among a set of organizations. A notable case in point is COMET, an international organization of metro (underground railway) operators that shares performance information, subject to the rule that it may not be published as a league table, to help each of the members to evaluate their relative performance and design policies accordingly (Glaister and Anderson 2005).

Using performance indicators as intelligence has numerous advantages. It avoids the ratchet and threshold effect problems associated with target systems. It avoids the compositional problems associated with ranking systems, and by avoiding a ranking format, the indeterminacy problems will also tend to be less serious. Because the consequences of the data collected are not easy for individuals and managers to predict, intelligence applications will also tend to be less subject to 'Goodhart's law' –

the proposition framed by a well-known economist in the 1980s that ‘Any observed statistical regularity will tend to collapse once pressure is placed on it for control purposes’ (Goodhart 1984: 94). Those problems will apply less to intelligence systems than to target or ranking systems.

But there are corresponding disadvantages. Where they lack transparency, intelligence applications of performance indicators may be more likely to be perceived as involving arbitrary judgement than target or ranking applications (though such judgements are of course far from absent in the latter cases). Moreover, such applications do not have the same potential as targets and ranking systems for externally motivating organizations and managers to concentrate their energies and improve their performance, in the form of the public shaming and praising effect of rankings and of the publicity and/or attention from higher authorities associated with target failure. They rely for their effectiveness either on intrinsic motivation by individuals and organizations to improve their performance or on the ability of external organizations to secure performance improvements by other means, such as the fear they instil common culture between the controllers and the controlled.

5. When to Use What? Dilemmas, Variable Returns and Contingencies

Having outlined three principal and common uses of performance indicators, what kind of theory might be advanced about when each use is to be preferred? Three possibilities are briefly considered here.

(a) The Dilemmas of Performance Indication

The first is that the choice among target, ranking and intelligence uses of performance indicators constitutes another of the many instances of administrative dilemmas (or poly-lemmas) – that is, situations in which any available choice involves advantages combined with unavoidable disadvantages. Such problems have been well-known in administrative science since Herbert Simon’s (1946) famous paper on the ‘proverbs of administration’ brought out the dilemmas involved in balancing conflicting desiderata in administrative design. So it could well be that the various available applications of performance indicators amount to yet another example of the same phenomenon. After all, as we have already noted, the design of target systems involve an inherent choice between the threshold effects prompted by across-the-board targets and the ratchet effects prompted by unit-specific targets, and the same goes to the choice between the output distortion prompted by ranking and target systems and the weaker motivation that may be prompted by intelligence systems.

Dilemmas by definition cannot be eliminated, so the task of the theorist amounts to identifying them and spelling them out explicitly, while the task of the policy designer is to be aware of them and craft policy creatively in the knowledge of their existence. Administrative and political systems typically deal with dilemmas by cycling among alternatives over time or by ‘clumsy’ systems (Thompson and Verweij 2006) that incorporate two or more approaches at the same time, without necessarily admitting the fact. And indeed examples of both cycling and clumsiness are not difficult to find in the applications of performance indicators as targets, rankings and intelligence.

(b) Variable or Diminishing Returns to Use

A second possibility, also familiar and simple, is that there may be variable returns associated with the three different uses of performance indicators considered here. Economists typically think of ‘diminishing returns’ as a phenomenon both of factors of production and of the process of consumption, so it would be natural to think of applications of performance indication as another instance of the law of diminishing returns.

Unlike administrative dilemmas in the pure sense, diminishing returns issues can be handled by theory or empirical study that identifies the shape of the returns curve and shows the point at which further investment in any one application exceeds the benefits it can secure (in principle taking into account also the opportunity costs of such increments of investment in terms of alternatives foregone). The difficulty with applying this apparently simple and obvious idea is that the costs of performance indication systems (like many other management systems) are rarely if ever collected and computed in a way that makes it possible to compare incremental returns against costs, even or perhaps especially in an age of supposedly evidence-based policy. Moreover, the shape of the returns associated with the various uses of performance indicators may be dramatically non-linear in some cases, meaning that an orthodox incremental perspective may not capture the point at which investment in such systems suddenly produces negative consequences for the system or society, as in the case, already mentioned, of those who see the collapse of the USSR as produced by the unintended effects of sixty years of management by targets.

Faced with diminishing returns, the task of the policy designer is to judge the optimum level of investment in any one application of performance indicators and as with administrative dilemmas, the implication may be to deal with the diminishing returns problem either by shifting the emphasis over time as diminishing or negative returns manifest themselves or by maintaining several applications at once.

(c) A Contingency Framework: ‘Horses for Courses’

The well-known principle of parsimony in scientific analysis states that we should only adopt more complex theories if simpler ones fail satisfactorily to account for the phenomena to be considered. Both the dilemma theory and the diminishing-returns theory could be argued to have strong elements of parsimony, and therefore perhaps deserve to be eliminated before any more complex account is considered. What neither of these simple theories can readily account for is why targets, rankings or intelligence work out differently in some organizational contexts than in others – for example with some organizations functioning as ‘honest triers’ and others as ‘gamers’. They do not readily explain why what might seem to be obviously rational systems of controlling public services are heavily resisted in some political cultures while they are eagerly embraced in others, as with the example given earlier of school league-tables in England and Belgium. Nor do they explain the circumstances in which one application of performance indicators is or should be preferred to the other two.

To the extent that more is going on in the application of performance indicators than can be accounted for by the presence of dilemmas or variable returns, a third possibility is to develop a form of ‘contingency theory’ to identify the circumstances

in which each application is most and least likely to be effective. After all, as mentioned earlier, the development of contingency theory was the way that administrative theory developed in the decades after Herbert Simon's identification of the previously unrecognized dilemmas underlying earlier general theories of administrative performance. That intellectual movement led to developments such as Henry Mintzberg's (1993) famous 'structure in fives' approach to organizational design. And indeed on the face of it, the various applications of performance indicators would seem to be ripe for the same sort of treatment.

Contingency theory developed in administrative science in the 1960s and 1970s to argue that there was no single all-purpose best way of structuring organizations, but that the most effective structure depended on the size, technology and environment of any one particular organization.³ (The approach weakened in the 1980s as rational-choice models of organization came into the ascendancy, but has subsequently re-emerged in a different form, in the guise of theories of cultural variability and 'clumsy' systems). Applied to the three different applications of performance indicators considered here, each of the three conventional variables used in 1960s-style contingency theory merits some attention, but 'environment' is the one that will be briefly developed here.

In the original development of contingency theory from the late 1950s, 'environment' was mainly conceived in terms of market stability or volatility, but can also be construed to denote organizations' task and cultural or political environments, and both of those aspects of environment seem to be applicable to the applications of performance indicators. As far as task environment is concerned, a simple form of contingency hypothesis is set out in Table 1. The suggestion embodied in that table is that the most appropriate use of targets comes when the intention is to concentrate the attention of individuals or organizations on improving performance in a limited number of areas. The most appropriate use of rankings comes when the intention is to encourage a set of producers to use whatever resources they have to improve their performance in defined ways without incurring the ratchet or threshold effects that accompany target systems. And the most appropriate use of intelligence comes when the intention is to develop learning capacity and diagnostic power rather than simply to stretch or sweat existing methods of production, and without incurring the output distortions that accompany target or ranking systems.

³ See for example Emery and Trist 1965; Blau and Schoenherr 1971; Pugh and Hickson 1976

Table 1: Task Contingencies and Three Applications of Performance Indicators

Intended effect	Use Performance Indicators as			Limits
Concentrating attention on improving performance in a limited number of areas	Targets			Ratchet effects threshold effects Output distortions
Sweating and stretching existing use of resources to improve performance		Rankings		Indeterminacy Volatility in composition Output distortions
Developing learning capacity and diagnostic power by adding knowledge about performance			Intelligence	Perception of arbitrariness, especially where application lacks transparency; Often rely on intrinsic motivation to improve

Source: Modified from Hood 2007: 101

That is a simple instrumental contingency hypothesis, relating the three possible uses of performance indicators to task environment, and it is only a little less parsimonious than the dilemma theory or the variable returns theory. But of course it assumes a culture-free world in which variations in attitudes and beliefs count for nothing in the development of performance systems. To relax that assumption, Table 2 offers a hypothesis about the way that different public service workgroup cultures might be expected to relate to the three applications of performance indicators discussed here. The cultural types come from the well-known grid-group cultural theory that is traceable back to Emile Durkheim's (1912) work on the elementary forms of the religious life, but was notably developed by Mary Douglas and her followers since the 1960s. Douglas and her followers argue that cultural variability tends to take the form of four conflicting worldviews, usually labelled as hierarchist, egalitarian, individualist and fatalist, and Table 2 offers a hypothesis about the general congruence to be expected between each of those worldviews in public service workgroup cultures and the three applications of performance indicators, together with a hypothesis about the most favoured applications (if any) for each worldview.

Table 2: Four Public Service Workgroup Cultures and Their Congruence with Three Applications of Measured Performance Indicators: A Hypothesis

Workgroup culture	Hierarchist	Egalitarian	Individualist	Fatalist
Application				
Targets	General congruence: medium to high Favoured applications: those that fit with official 'war effort' goals	General congruence: medium to high Favoured applications: those that fit with equalizing goals	General congruence: variable Favoured applications: those that lead to individual advancement	General congruence: fairly low Favoured applications: none. Sceptical or resigned acceptance in all cases
Rankings	General congruence: medium to low Favoured applications: those that maintain established pecking orders and administered by authority figures	General congruence: generally low Favoured applications: those that upset or challenge established pecking orders e.g. value added tables for schools	General congruence: generally high Favoured applications: those that involve individual as well as organizational rankings (e.g. university research assessments)	General congruence: fairly low Favoured applications: none. Likely to be seen as a lottery in all cases
Intelligence	General congruence: generally high Favoured applications: those that involve recognized 'wise heads' and differentiated security levels	General congruence: medium to low Favoured applications: those that involve sharing of data for workgroup or enclave rather than top-down uses	General congruence: generally low Favoured applications: those that give opportunities for individual access to sources of influence	General congruence: fairly low Favoured applications: few. Likely to be expected as a feature of rule but to have low motivational impact

The hypothesis suggested here is that fatalist cultures are unlikely to offer strong resistance to any of those applications, but equally that none of those applications are likely to have strong motivational effects in such a cultural environment. By contrast, hierarchist cultures are expected to be able to respond to all of the three applications, and are likely to accept more readily a 'wise heads' view of the uses of intelligence than egalitarians or individualists. Egalitarian cultures are likely to be more resistant to rankings than the other two applications, and are likely to prefer collectivized targets and enclave-bound intelligence, while for individualist cultures rankings are likely to be the preferred application of performance indicators (particularly where they can be sheeted back to individual rather than group performance). Such an analysis, only sketched out briefly in Table 2, might help to account both for variable reactions and use of the three applications in different contexts – for example, England's love affair with performance rankings for schools and hospitals in the 2000s but the utter rejection of such applications in Belgium – and for the sort of reasons that prompt the application of performance indicators as targets in some cases, as rankings in others and as intelligence in yet others.

6. Conclusion

Three conclusions can be briefly drawn from this analysis. One is that the use of public service performance indications for 'intelligence' is remarkably little documented and theorized, in spite of the massive practico-descriptive literature on public sector performance indicators that has developed in recent decades. Second, up to this point we lack an overall theory – whether of an instrumental or of a social and historical kind – of application of performance indicators as targets, rankings or intelligence, although the scope and limits of at least the first two applications have been the subject of considerable scholarship. And third, if such a theory needs to be more than the identification of dilemmas or diminishing returns to use, some sort of contingency approach of the kind set out very sketchily here seems unavoidable even with due attention to the parsimony principle.

References

- Bain, J, Keren, J, Miller, J and Thornton, J (1987) 'The Ratchet, Tautness and Managerial Behaviour in Soviet-Type Economies' European Economic Review 31: 1173-1201
- Barber, M (2007) Instruction to Deliver: Tony Blair, Public Services and the Challenge of Achieving Targets, London, Politico's
- Besançon, M (2003) Good Governance Rankings: The Art of Measurement, WPF Reports No. 36, Cambridge, MA, World Peace Foundation
- Blau, P M (1955) The Dynamics of Bureaucracy, Chicago, Chicago University Press
- Blau, PM and Schoenherr, R A (1971) The Structure of Organizations, New York, Basic Books
- Braguinsky S and Yavlinksi, G (2000) Incentives and Institutions: The Transition to a Market Economy in Russia, Princeton, Princeton University Press
- Brown, P Miller, J and Thornton, J (1994) 'The Ratchet Effect and the Coordination of Production in the Absence of Rent Extraction' Economica 61 (241): 93-114
- Deming, W E (2000) The New Economics for Industry, Government, Education, 2nd edition, Boston, MIT Press

- Dobb, M (1970) Socialist Planning: Some Problems, London, Lawrence and Wishart
- Dunsire, A (1978) The Execution Process, Vol 2: Control in a Bureaucracy, Oxford, Martin Robertson
- Elmore, R (2000)
- Emery, F E and Trist, E L (1965) 'The Causal Texture of Organisational Environments' Human Relations 18
- Glaister, S and Anderson, R (2005) 'Transport Performance and the Data Clubs Approach' Where Does Britain Rank? - International Public Services Rankings Conference presentation (accessed June 2007)
http://www.publicservices.ac.uk/Events/Rankings_conference/Rankingswebpage.asp
- Goodhart CAE (1984) Monetary Theory and Practice: The UK experience, London, Macmillan
- Gormley W T and Weimer, D L (1999) Organizational Report Cards, Cambridge, MA, Harvard University Press
- Hall, R H (1972) Organizations: Structure and Process, Englewood Cliffs, NJ, Prentice-Hall
- Hood, C (2007) 'Public Management by Numbers: Why does it Vary? Where Has it Come From? Where are the Gaps and the Puzzles?' Public Money and Management 27 (2): 95-102
- Hood, C, Dixon, R and Beeston, C (2007) Rating the Rankings: Assessing International Rankings of Public Service Performance' paper presented to IPMN Workshop, 'Ratings and Rankings of Public Services', Oxford, 7-9 August 2007
- Hume, L J (1981) Bentham on Bureaucracy, Cambridge, Cambridge University Press
- Jacobs, R and Goddard, M (2007) 'How Do Performance Indicators Add Up? An Examination of Composite Indicators in Public Services,' Public Money and Management 27 (2): 103-110
- Kornai, J (1992) The Socialist System: The Political Economy of Communism, Oxford, Clarendon
- Lazear, E and Rosen, S (1981) 'Rank-Order Tournaments as Optimum Labor Contracts' Journal of Political Economy 89: 841-64
- Mintzberg (1993) Structure in Fives, Englewood Cliffs, NJ, Prentice-Hall
- Misra, B B (1970) The Administrative History of India, 1834-1947: General Administration, Oxford, Oxford University Press
- Pugh, D S and Hickson, D J (1976) Organisation Structure in its Context: The Aston Programme 1, Farnborough, Saxon House
- Simon, H A (1946) 'The Proverbs of Administration' Public Administration Review 6 (1): 53-67
- Taylor, F W (1911) The Principles of Scientific Management, New York, Harper
- Verweij, M and Thompson, M eds (2006) Clumsy Solutions for a Complex World: Governance, Politics and Plural Perceptions, London, Palgrave